What is the difference between controlling for mean versus median income in analyses of income inequality?

T A Blakely, I Kawachi

Department of Health and Social Behavior at the Harvard Center for Society and Health, Harvard School of Public Health, Harvard University, Boston, USA
T A Blakely
I Kawachi

Department of Public Health, Wellington School of Medicine, University of Otago, Wellington, New Zealand
T A Blakely

Correspondence to:
Dr Blakely, Department of Public Health, Wellington School of Medicine, University of Otago, PO Box 7343, Wellington, New Zealand
(tblakely@wnmeds.ac.nz)

Accepted for publication 13 November 2000

It is routine to control for "average" income when assessing the independent effect of income inequality on health, but authors have used different measures, for example, percentage poverty,[1] per capita or mean income,[2] and median income.[3 4] However, as the distribution of income in a population is always positively skewed (that is, a long thin tail for the few with high incomes), the median income and percentage poverty are necessarily correlated with any measure of income inequality. For example, for a given total income of $1 billion in population of 100 000 people, the average income is the same ($10 000) regardless of how that income is distributed. Assume the distribution of income is log-normal (that is, the proportion of people at each income level (y axis) has a normal distribution plotted against the log of that income (x axis)). Figure 1 plots the cumulative proportion of the population with up to the given income on the x axis, for three scenarios: low income inequality (SD log income = 0.4); medium income inequality (SD 0.6) and high income inequality (SD 0.8). As the total income is fixed (and hence the average income is fixed), the median income must differ under each scenario and is given by the intercept of each curve with the gridline for the 0.5 cumulative proportion: $8777, $7942, and $6904 for the low, medium, and high income inequality scenarios, respectively. Note that as the income inequality increases, the median income decreases—a necessary association for any positively skewed shape of the income distribution. Thus, median income (and likewise percentage poverty) reflects both

the total income and the income distribution—however, mean income reflects only the total income. If in analyses of the association of income inequality we wish to control for average wealth as a possible contextual confounder, it would therefore appear preferable to use mean income rather than median income or percentage poverty.

In this paper, we empirically test for a difference between using mean and median household income to control the association of income inequality at the metropolitan area (MA) in the US.

Methods

Mean and median household income for MAs were taken from the 1990 census for 232 MAs that were identified in the 1996 and 1998 March Current Population Surveys (CPS; administrative boundaries the same on each dataset). Gini coefficients (measure of income inequality) were calculated for these MAs using the 1990 census data with the income distribution software developed by Ed Welniak (1988, US Census Bureau). Pearson correlations were conducted among the three MA level variables (mean income, median income, and Gini). The association of income inequality with fair/poor self rated health was determined using a multilevel analysis with Proc Glimmix in SAS . The individual level data were the 185 479 respondents to the CPS in 1996 and 1998 (combined), with categorical variables created for age, race, sex, household income, and fair/poor self rated health.

Results

Across the 232 MAs the mean Gini coefficient was 0.417 (SD 0.025), and using the mean and mean (SD) as cut off points, 34, 71, 95, and 32 MAs were assigned to high, medium-high, medium-low and low categories of income inequality. The mean and median household incomes for the 232 MAs were themselves distributed with means of $37 713 (SD $7890) and $30 317 (SD $6207), respectively. One MA (Stamford CT) was an obvious outlier with a median and mean household income 8% and 50%, respectively, greater than the next wealthiest MA, and the 10th highest Gini. Correlation coefficients excluding Stamford are reported first followed by those including Stamford—all are highly statistically significant (p<0.0001).

The mean and median household incomes were highly correlated, *r*=0.97 and 0.95. Yet, the correlation coefficient of the median household income with Gini (*r* = −0.59 and

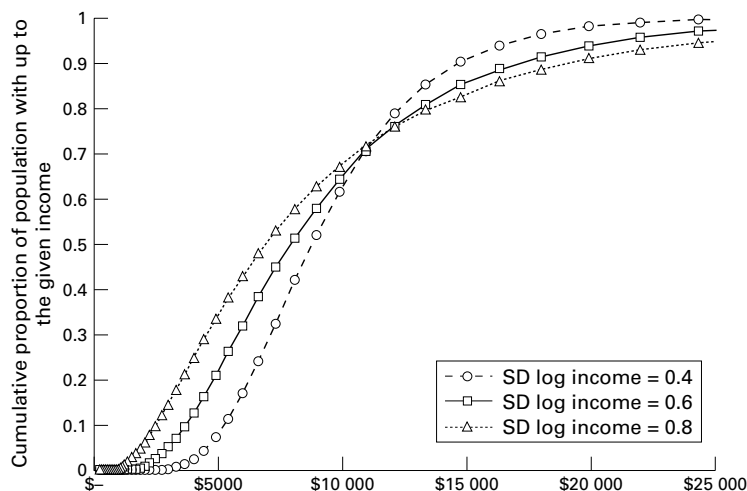*Figure 1   The cumulative proportion of total income by actual income when the average income is fixed at $10 000 and distributed log-normally, but the standard deviation of the log-normal distribution varies (that is, income inequality varies).*

Table 1 Odds ratios of fair/poor health by MA level income inequality, mean and median household income, for CPS sample with MA Gini assigned (n=185 479) ★

	Gini only	Gini and mean income	Gini and median income
Income inequality (MAs)			
High (n=34)	1.20 (1.04, 1.38)	1.09 (0.94, 1.25)	1.02 (0.88, 1.20)
Medium-high (n=71)	1.07 (0.95, 1.21)	1.00 (0.88, 1.12)	0.97 (0.85, 1.10)
Medium-low (n=95)	1.02 (0.91, 1.15)	0.97 (0.87, 1.09)	0.96 (0.85, 1.08)
Low† (n=32)	1.00	1.00	1.00
Log of mean/median household income		0.65 (0.53, 0.80)	0.65 (0.52, 0.80)

★The odds ratios are from logistic regression models for fair/poor self rated health that included the variables in each column, and individual level sex, age, race and equivalised household income. A random intercept was specified at the MA level in each model. Analyses were weighted using the CPS "final weight". †Reference category.

−0.52) were substantially greater than that of the mean household income with Gini ($r = -0.41$ and -0.29). In multi-level analyses including Stamford (excluding Stamford made no substantive difference) there was a small association of income inequality with fair/poor self rated health when controlling for MA level mean household income (odds ratio 1.09 for high compared with low income inequality MAs, 95% confidence intervals 0.94, 1.25; see table 1). Controlling for MA level median household income the income inequality association was further reduced to the null (odds ratio 1.02 (0.88, 1.20)).

Discussion

While it is debatable whether income inequality at the MA level is associated with self rated health at all (a more comprehensive paper is being published elsewhere[5]), the point we wish to make here is that the specification of "average" income seems to affect the size of the income inequality association with self rated health. Thus, not only is there a theoretical preference for using mean rather than median income, but using median household income seems to "over-control" the association of income inequality compared with using mean household income. The fact that the estimated income inequality effect size is small controlling for either mean or median income (with both confidence intervals overlapping 1.0) does not, we believe, invalidate our conclusion. Firstly, just as the decision to include confounders in a model should not be based on statistical tests, but rather the impact on the point estimate of

the effect measure, it is likewise inappropriate to use statistical criteria to "test" for a difference between mean and median income. Moreover, both the mean and median variables are derived from the same underlying income data, and thus varying impacts of the income inequality effect measure are not the result of sampling error. Secondly, we choose to present results controlling for individual level income—models that exclude individual level income have higher odds ratios for income inequality (confidence intervals excluding 1.0), and there was a similar relative change in the point estimate between controlling for mean versus median income. Accordingly, we recommend that researchers use mean income in future analyses.

1 Kennedy B, Kawachi I, Prothrow-Stith D. Income distribution and mortality: cross sectional ecological of the Robin Hood index in the United States. *BMJ* 1996;**312**:1004–7.
2 Lynch J, Kaplan G, Pamuk E, *et al*. Income inequality and mortality in metropolitan areas of the United States. *Am J Public Health* 1998;**88**:1074–80.
3 Kaplan G, Pamuk E, Lynch J, *et al*. Inequality in income and mortality in the United States: analysis of mortality and potential pathways. *BMJ* 1996;**312**:999–1003.
4 Soobader M-J, LeClere F. Aggregation and the measurement of income inequality: effects on morbidity. *Soc Sci Med* 1999;**48**:733–44.
5 Blakely T, Lochner K, Kawachi I. Metropolitan area income inequality and self-rated health: a multi-level study. *Soc Sci Med* (in press).